

Г. Н. ЖОЛТКЕВИЧ, д-р техн. наук,
АХМАД ЮСЕФ ИБРАХИМ ИБРАХИМ, аспирант, ХНУ им. В.Н. Каразина

МОДЕЛЬ ХРАНЕНИЯ ОБРАЗЦОВ ПОЛУСХЕМ СРЕДСТВАМИ РЕЛЯЦИОННЫХ БАЗ ДАННЫХ

В статті розглянута задача забезпечення якості інформації в інформаційних системах за рахунок формалізації структури інформації, якою обмінюються система та її оператори. Структурування інформації, як метод забезпечення її коректності, пропонується проводити на основі єдиної мета моделі – теорії напівсхем. Запропонована концептуальна модель представлення зразків напівсхем в інформаційній системі.

Качество информации является одним из определяющих факторов качества информационных управляющих систем. Актуальность проблемы обеспечения качества информации в информационных управляющих системах связана, в первую очередь, с тем, что принимаемые на основании недостоверной, неактуальной, неточной информации управленческие решения могут привести к серьезным проблемам для объекта управления. Рассматривая проблему обеспечения качества информации в информационных системах, исследователи выделяют различные аспекты ее решения [1 – 5]. В качестве одного из факторов повышения качества информации является формализация структуры документов и создание интеллектуальных программных средств поддержки оборота формализованных документов. Введение формализованных документов должно базироваться на строгих математических моделях, отражающих концептуальный (структурно-логический) аспект предметной области. Авторами в настоящей работе в основу построения таких моделей положена общая алгебраическая метамодель (теория полусхем), впервые предложенная Жолткевичем Г. Н. и Семеновым Т. В. в работе [6]. Дальнейшее развитие этот аппарат получил благодаря работе [7], авторами которой удалось представить общую метамодель средствами реляционной алгебры и, тем самым, обеспечить ее технологическую поддержку. Несмотря на то, что теория полусхем оперирует с экземплярами (образцами [6]) понятий, однако исследование методов их обработки остаются за ее рамками. В связи с этим представляется перспективным расширение метамодели в этом направлении. Первый шаг сделан в работе [8], где показано, что каждый образец понятия полусхемы изоморфен некоторому маркированному графу. Несмотря на теоретическую важность этого результата, проясняющего структуру образцов понятий в теории полусхем, практическое его использование требует технологического обеспечения обработки и хранения маркированных деревьев, представляющих образцы понятий.

Целью настоящей работы является исследование возможности использования реляционной модели данных для организации хранения образцов полусхемы.

В настоящей работе используется система обозначений и определений работ [6, 7], которая приводится ниже.

Для пары множеств X , Y обозначим

$M_+(X, Y)$ – множество частичных, хотя бы где-то определенных, отображений из X в Y ;

$\text{dom}(f)$ – область определения отображения $f \in M_+(X, Y)$;

$\text{im}(f)$ – область значения отображения $f \in M_+(X, Y)$;

xR – подмножество $\{y \in Y \mid (x, y) \in R\}$ для бинарного отношения R между множествами X и Y ;

X^* – множество конечных последовательностей элементов множества X , включающее пустую последовательность, которая всегда обозначается символом ε ;

$h(x_1 \dots x_k) = x_1$ – для любой непустой последовательности элементов множества X ;

$t(x_1 x_2 \dots x_k) = x_2 \dots x_k$ – для любой непустой последовательности элементов множества X .

Полусхемой предметной области назовем тройку $S = (N, R, D)$, где N , R – конечные множества понятий и ролей соответственно, $D \subset N \times M_+(R, N)$, для которой выполняется условие:

для $n \in N$, $f, g \in M_+(R, N)$, $r \in R$ таких, что $(n, f) \in D$, $(n, g) \in D$ и $r \in \text{dom}(f) \cap \text{dom}(g)$, верно $f(r) = g(r)$.

Для полусхемы $S = (N, R, D)$ понятие $n \in N$ называется **базовым**, если для всякого $f \in M_+(R, N)$ выполняется $(n, f) \notin D$. Множество базовых понятий полусхемы будем обозначать N_0 .

Пусть $S = (N, R, D)$ является полусхемой и для некоторого $n \in N$ существует $f \in M_+(R, N)$ такое, что $(n, f) \in D$. Тогда будем говорить, что для понятия n задан **вариант определения** f .

В работе [7] показано, что полусхема может быть однозначно представлена реляционной базой данной с двумя отношениями. Концептуальная модель базы данных в нотации языка UML представлена на рис. 1. Соответствующие этой модели семантические ограничения приведены ниже на языке OCL (Object Constraint Language) [9].

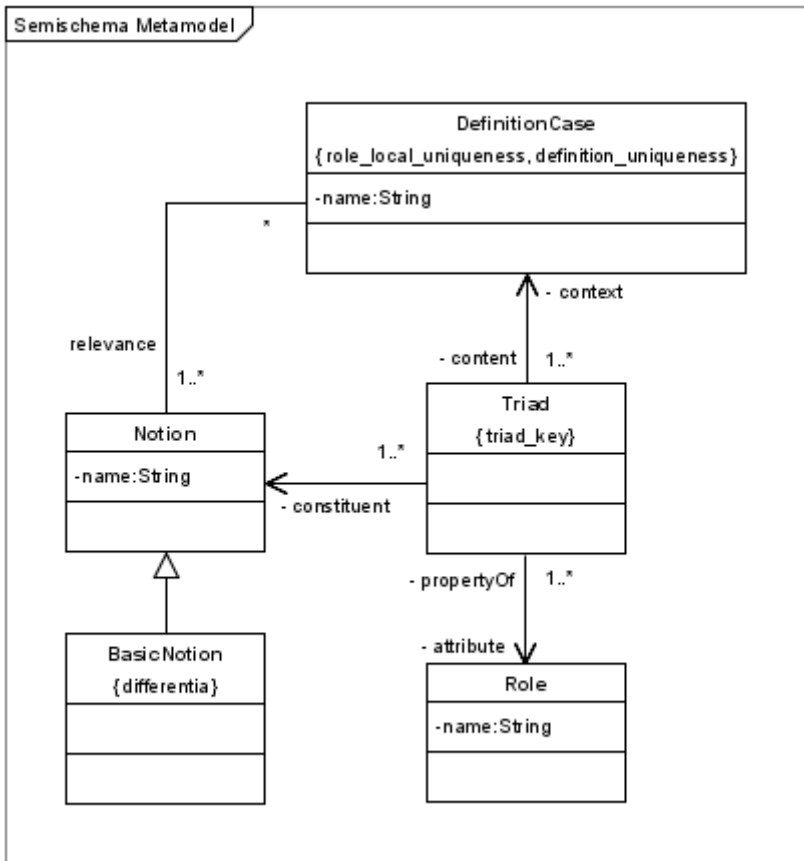


Рис. 3. UML метамодель полусхем.

package SemischemaMetamodel

context BasicNotion **inv** differentia :
self.relevance->size() = 0

context Triad **inv** triad_key :
self.attribute.propertyOf->intersection(
self.context.content) = Set{self}

context DefinitionCase **inv** definition_uniqueness :
DefinitionCase.allInstances()->select(f : DefinitionCase |
f.content = self.content) = Set{self}

context DefinitionCase **inv** role_local_uniqueness :

```

DefinitionCase.allInstances()->forAll(f |
  DefinitionCase.allInstances()->forAll(g |
    f <> g and f.relevance->intersection(
      g.relevance)->notEmpty() and
    forAll(r : Role |
      f.content.attribute->intersection(
        g.content.attribute)-
>includes(r) implies
      f.content->select(t : Triad |
        t.attribute = r).constituent =
      g.content->select(t : Triad |
        t.attribute = r).constituent))))

```

endpackage

В работе [8] определено маркированное дерево.

Под деревом мы будем понимать ориентированный граф

$$G = (\mathbf{V}, \mathbf{E}, \text{beg} : \mathbf{E} \rightarrow \mathbf{V}, \text{end} : \mathbf{E} \rightarrow \mathbf{V}),$$

обладающий следующими свойствами:

1. множество $\mathbf{V} - \text{end}(\mathbf{E})$ состоит из одного элемента, который обозначается через root и называется корнем дерева;
2. для каждого $v \in \mathbf{V}$, отличного от root , существует единственный путь, началом которого является root , а концом v .

Маркированным деревом называется восьмерка

$$T = (\mathbf{V}, \mathbf{E}, M_{\mathbf{V}}, M_{\mathbf{E}}, \text{beg} : \mathbf{E} \rightarrow \mathbf{V}, \text{end} : \mathbf{E} \rightarrow \mathbf{V}, \\ m_{\mathbf{V}} : \mathbf{V} \rightarrow M_{\mathbf{V}}, m_{\mathbf{E}} : \mathbf{E} \rightarrow M_{\mathbf{E}}),$$

для которой $(\mathbf{V}, \mathbf{E}, \text{beg} : \mathbf{E} \rightarrow \mathbf{V}, \text{end} : \mathbf{E} \rightarrow \mathbf{V})$ является деревом.

Как показано в работе [8], образцы понятий, определенные в [6] являются специальным образом маркированными деревьями.

Для полусхемы $S = (\mathbf{N}, \mathbf{R}, \mathbf{D})$ маркированное дерево

$$T = (\mathbf{V}, \mathbf{E}, \mathbf{D} \cup \mathbf{N}_0, \mathbf{R}, \text{beg} : \mathbf{E} \rightarrow \mathbf{V}, \text{end} : \mathbf{E} \rightarrow \mathbf{V}, \\ m_{\mathbf{V}} : \mathbf{V} \rightarrow \mathbf{D} \cup \mathbf{N}_0, m_{\mathbf{E}} : \mathbf{E} \rightarrow \mathbf{R}),$$

удовлетворяющее следующим условиям:

1. $m_{\mathbf{V}}(v) \in \mathbf{N}_0$ тогда и только тогда, когда v – лист дерева;
2. если $m_{\mathbf{V}}(v) = (n, f) \in \mathbf{D}$, то $m_{\mathbf{E}}$ устанавливает взаимно однозначное соответствие между $\{e \in \mathbf{E} \mid \text{beg}(e) = v\}$ и $\text{dom}(f)$;

3. если $v, v' \in V$, $e \in E$, $\text{beg}(e) = v$, $\text{end}(e) = v'$, $m_v(v) = (n, f)$, а $m_v(v') = (n', g)$ (для некоторого g) или $m_v(v') = n'$, и $m_E(e) = r$, то $f(r) = n'$,

будем называть образцом понятия n_0 , где $m_v(\text{root}) = (n_0, f_0)$ для некоторого f_0 .

На рис. 2 представлена концептуальная модель образца полусхемы и далее приведены ограничения целостности на языке OCL.

package PatternModel

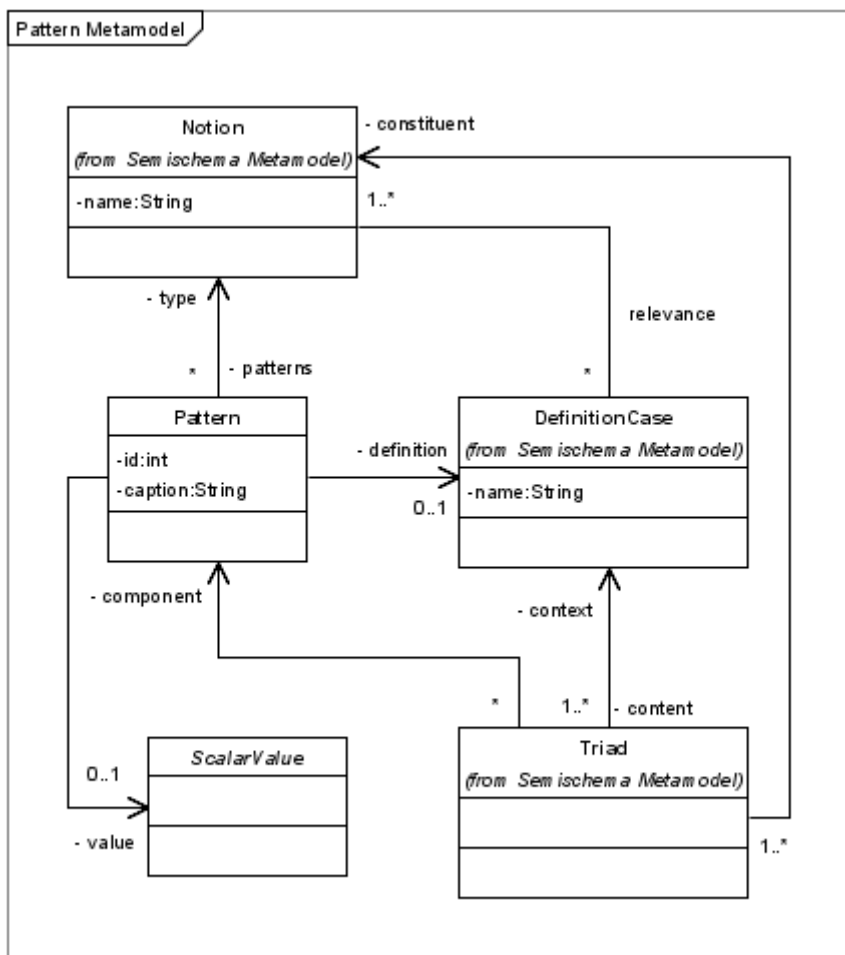


Рис. 4. UML метамодель образца полусхемы

```

context Pattern inv basic_or_complex :
    self.type.occlIsType(BasicNotion) implies
        self.definition->isEmpty() and
        self.value->notEmpty()

and
not self.type.occlIsType(BasicNotion) implies
    self.definition->notEmpty() and
    self.value->isEmpty()

context Pattern inv integrity :
    self.type.relevance->includes(self.definition)

```

endpackage

Таким образом, в настоящей работе

1. показано, что для хранения образцов полусхемы можно использовать реляционную модель данных, концептуальная модель которой представлена на рис. 2;
2. выявлены семантические ограничения для предложенной модели.

В дальнейшем необходимо построить аксиоматическую модель реляционной базы данных для хранения образцов полусхемы и разработать методы проверки корректности хранилища образцов полусхемы, базирующиеся на реляционном исчислении.

Список литературы: 1. *Rose F.* The Economics, Concept, and Design of Information Intermediaries. – Berlin: Springer, 1999. – 266 p. 2. *Аксенов Е.* Качество информации: от очистки данных – к модели предприятия. – PCWEEK. Корпоративные системы. – №36(354), 2002. – <http://www.pcweek.ru/Year2002/N36/CP1251/CorporationSystems/chapt1.htm>. 3. *Eppler M. J.* Managing Information Quality. – Berlin: Springer, 2003. – 302 p. 4. *Sy B. K., Gupta A.K.* Information-Statistical Data Mining: Warehouse Integration with Examples of Oracle Basics. – The International Series in Engineering and Computer Science. – Vol. 757, 2004. – 312 p. 5. *Bacharach M., Board O.* The quality of information in electronic groups. – Netnomics. – № 4, 2002. – Pp. 73 – 97. 6. *Жолткевич Г. Н., Семенова Т. В.* К проблеме формализации концептуального моделирования информационных систем. – Вісник Харк. нац. ун-та. Серія „Математичне моделювання. Інформаційні технології. Автоматизовані системи управління”. – № 605(2), 2003. – С. 33 – 42. 7. *Жолткевич Г. Н., Семенова Т. В., Федорченко К. А.* Представление полусхем предметных областей информационных систем средствами реляционных баз данных. – Вісник Харк. нац. ун-та. Серія „Математичне моделювання. Інформаційні технології. Автоматизовані системи управління”. – № 629(3), 2004. – С. 11 – 24. 8. *Жолткевич Г. Н., Ибрахим Ахмад Юсеф Ибрахим.* О возможности представления образцов понятий полусхем маркированными деревьями. – X.: Системи обробки інформації. – Вип. 2 (51), 2006. – С. 20 – 26. 9. OMG Unified Modelling Language Specification. Version 1.5. – An Adopted Formal Specification of the Object Management Group, Inc. – 2003. – 736 p.

Поступила в редколлегию 02.04.06